1-2019

# Algorithm mismatch in spatial steganalysis

Stephanie Reinders
*Iowa State University*, srein@iastate.edu

Jennifer Newman
*Iowa State University*, jlnewman@iastate.edu

Li Lin
*Iowa State University*

Yong Guan
*Iowa State University*, guan@iastate.edu

Min Wu
*University of Maryland at College Park*

# Algorithm mismatch in spatial steganalysis

## Disciplines

Forensic Science and Technology

## Comments

Posted with permission of CSAFE.

*Electronic Imaging 2019*

csafe

Center for Statistics and Applications in Forensic Evidence

www.forensicstats.org

# *Algorithm Mismatch in Spatial Steganalysis*

Stephanie Reinders, Department of Mathematics, Iowa State University
Dr. Jennifer Newman, Department of Mathematics, Iowa State University
Li Lin, Department of Mathematics, Iowa State University
Dr. Yong Guan, Department of Electrical and Computer Engineering, Iowa State University
Dr. Min Wu, Department of Electrical and Computer Engineering, University of Maryland

# Overview

- Introduction
- Algorithm Mismatch Approach to Blind Detection
- Algorithm Mismatch Results
- Conclusion

# Steganography and Steganalysis

Cover

+

Payload text


=

Steganography: "*covered writing*" (Greek)

Steganalysis: detecting hidden messages

Stego



csafe
Center for Statistics and Applications in Forensic Evidence

# Algorithm Mismatch in Blind Detection

- Motivation
  - Our ultimate goal is to create a uncomplicated, non-data-intensive framework for blind steganalysis
  - The framework consists of a single binary classifier with single known embedding algorithm that will detect stegos with other unknown embedding algorithms
  - Addresses the real-world scenario where training data and processing time are limited
- *Algorithm Mismatch* is the case where a classifier is trained on one stego embedding algorithm and tested others

# Previous Blind Detection Approaches

- Multi-classifier approach
  - Set of (n choose 2) binary classifiers predicts cover or stego and the embedding algorithm [Pevny and Fridrich 2008]
- Domain adaptation approach
  - Domain adaptation in DCT domain to predict cover or stego [Kong et. al. 2016]
- Other approaches
  - One-class classifier trained on covers [Pevny and Fridrich 2008]
  - Binary classifier trained on covers and collection of stego algorithms predicts cover or stego [Pevny and Fridrich 2008]
  - Clustering predicts guilty actors in pooled steganalysis [Ker and Pevny 2011]

# Algorithm Mismatch in Blind Detection

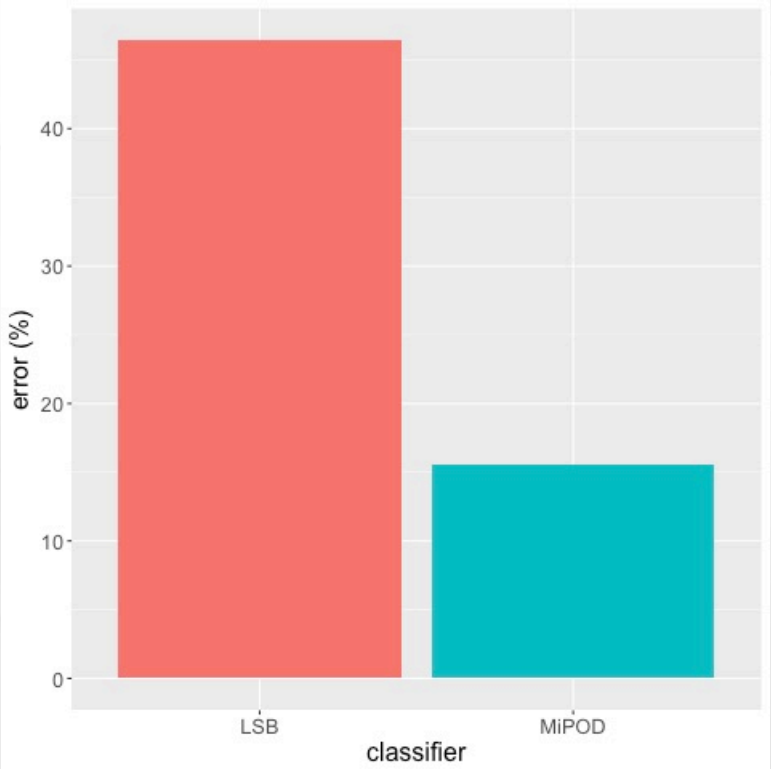
- *Algorithm Mismatch* is the case where a classifier is trained on one stego embedding algorithm and tested other algorithms
- We consider the case where spatial embedding algorithms make changes in the LSB plane only
- We adjust the Ensemble Classifier with Spatial Rich Model features
- 4 embedding algorithms: LSB matching, MiPOD, S-UNIWARD, and WOW
- Uncomplicated, non-data-intensive approach to algorithm mismatch in blind steganalysis

# Datasets Used

- Dataset 1: BOSSbase
  - 10,000 RAW images from 7 digital still cameras. Converted to TIFF (Photoshop). Center-cropped 512x512 images, converted to grayscale and saved as PNG (Matlab)
  - 4 stego algorithms: LSB matching, MiPOD, S-UNIWARD, WOW
- Dataset 2: StegoAppDB – Forensic Image Database
  - 1,927 TIFF auto-exposure images from 6 iPhone devices. Cropped into 5 disjoint 512x512 images, converted to grayscale and saved as PNG (Matlab) for a total of 9,635 covers
  - 4 stego algorithms: LSB matching, MiPOD, S-UNIWARD, WOW

# Motivating Example

Average detection error on BOSSbase MiPOD data



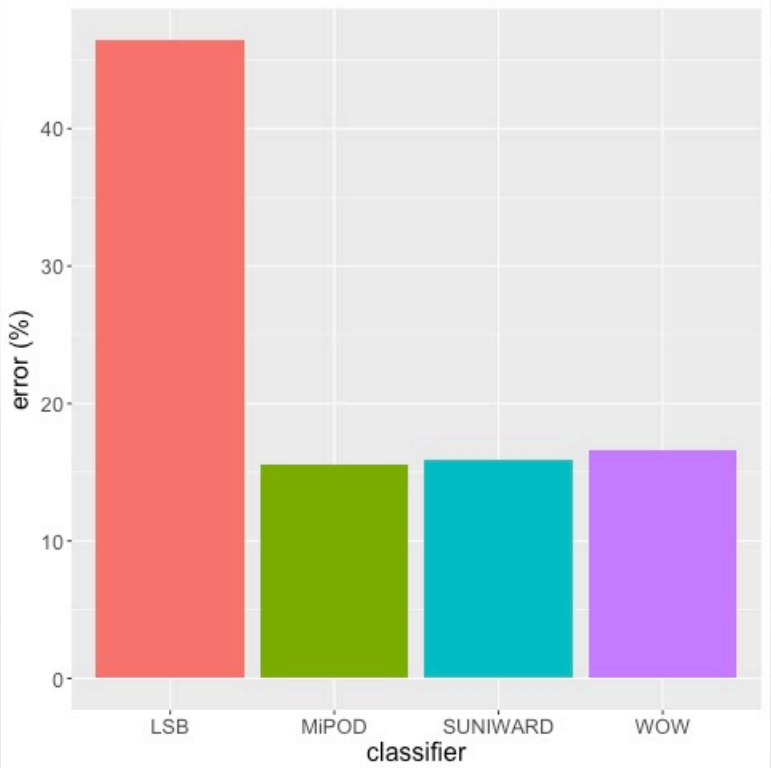

- Test data is BOSSbase covers and MiPOD with 10% embedding rate
- "best-case" classifier is a MiPOD trained classifier
- MiPOD classifier achieves 16% error rate
- LSB trained classifier has error rate close to random guessing



csafe Center for Statistics and Applications in Forensic Evidence

# Motivating Example

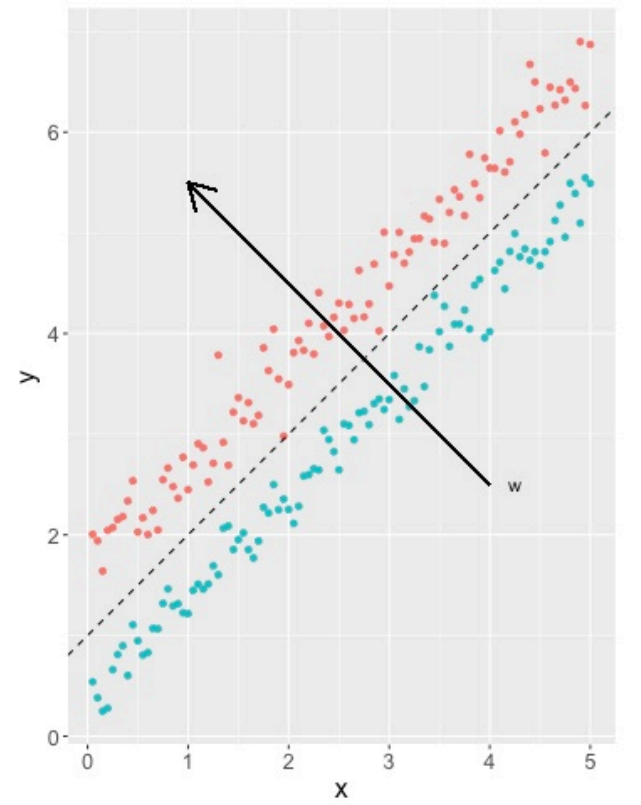
Average detection error on BOSSbase MiPOD data






csafe
Center for Statistics and Applications in Forensic Evidence

# Fischer Linear Discriminant

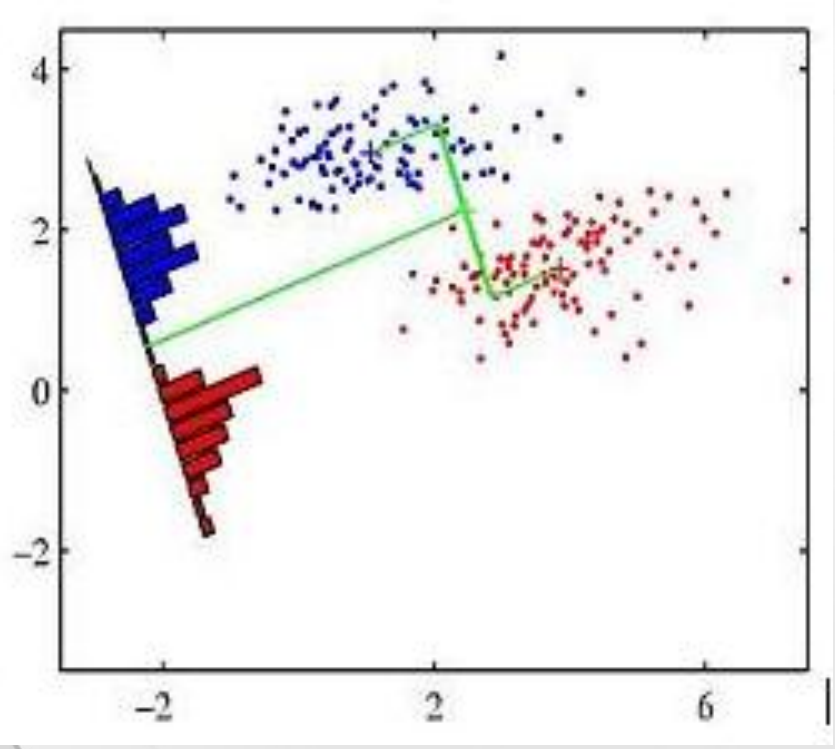



- The Ensemble Classifier consists of a set of base learners
- Each base learner is a Fischer Linear Discriminant
  - An FLD finds the vector $w$ that maximizes the between-class variance and minimizes the within-class variance
  - $w$ is the normal vector for the decision hyperplane $b$, also called the threshold
  - Test image $x$ is projected onto $w$

    $$g(x) = w^T x$$
  - The class is predicted based on which side of $b$ the projection lies

    $$\begin{cases} g(x) > b, & x \text{ is stego} \\ g(x) < b, & x \text{ is cover} \end{cases}$$
  - The *standard* threshold $b$ is chosen to minimize the False Alarm Rate and Missed Detection Rate

# Fischer Linear Discriminant




- The Ensemble Classifier consists of a set of base learners
- Each base learner is a Fischer Linear Discriminant
  - An FLD finds the vector $w$ that maximizes the between-class variance and minimizes the within-class variance
  - $w$ is the normal vector for the decision hyperplane $b$, also called the threshold
  - Test image $x$ is projected onto $w$
    $$g(x) = w^T x$$
  - The class is predicted based on which side of $b$ the projection lies
    $$\begin{cases} g(x) > b, & x \text{ is stego} \\ g(x) < b, & x \text{ is cover} \end{cases}$$
  - The *standard* threshold $b$ is chosen to minimize the False Alarm Rate and Missed Detection Rate

# Adjusting the Fisher Linear Discriminant

- When training on LSB data, we adjust the decision threshold

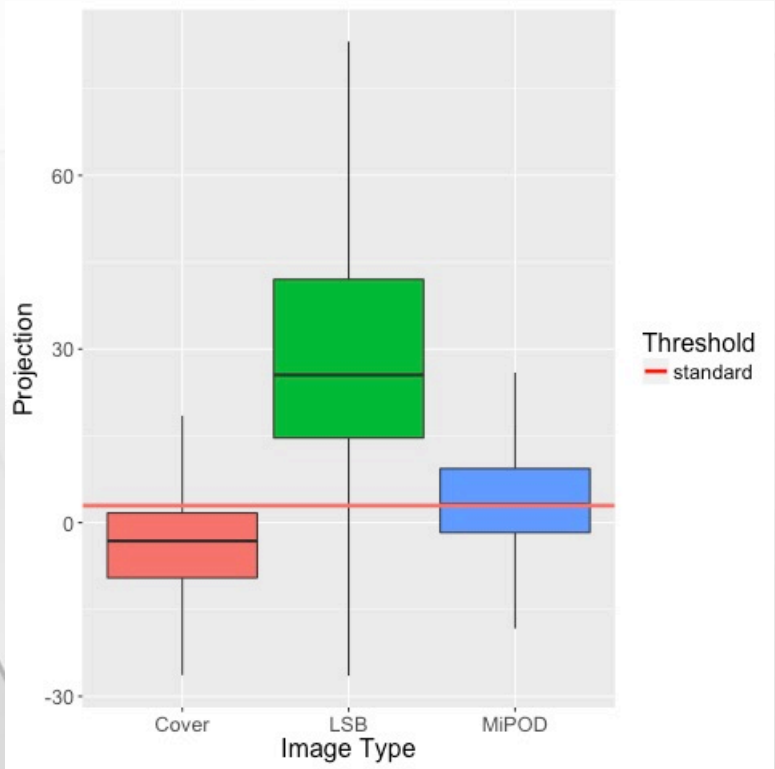
$$b_{adj} = b - \lambda c$$

  - $c$ is the standard deviation of the projections of the training data
  - $\lambda$ is a tuning parameter
  - Experimentally, we found the parameter $\lambda = 0.75$ to give good overall detection results when testing on MiPOD and S-UNIWARD
- Test image $x$ is projected onto $w$. The class is predicted based on which side of $b_{adj}$ the projection lies

$$\begin{cases} g(x) = w^T x > b_{adj} & x \text{ is stego} \\ g(x) = w^T x < b_{adj} & x \text{ is cover} \end{cases}$$

# Why Does Algorithm Mismatch Work?

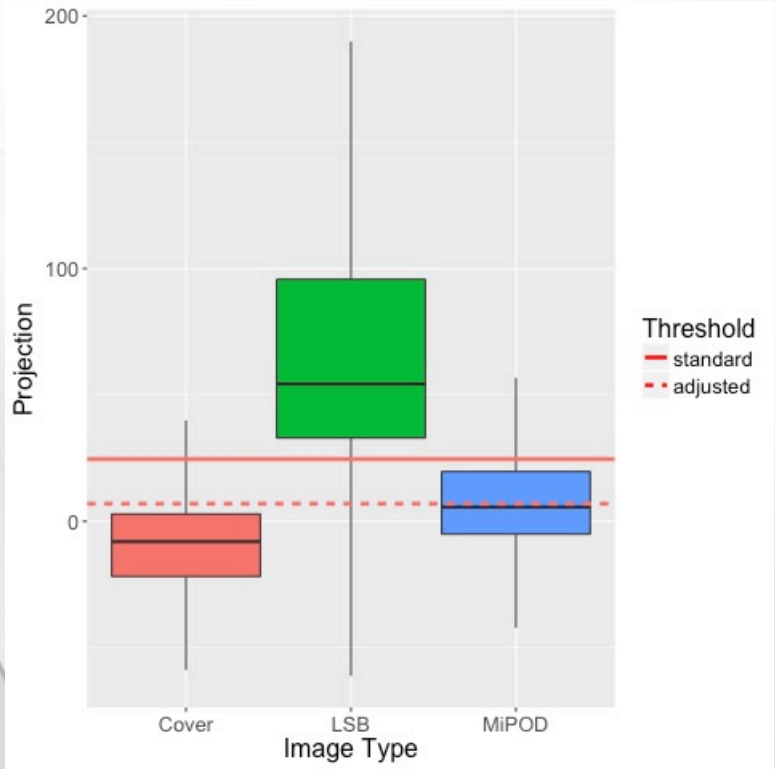
FLD projections of test image features onto normal vector of decision hyperplane




- Best-case classifier - MiPOD trained ensemble classifier with standard threshold
- 10% embedding rate
- 5 repetitions of 10-fold-cross validation
- Each test image projected onto normal vector of decision hyperplane
- Threshold is median across all base learners
- Outliers not shown (~12%)

# Why Does Algorithm Mismatch Work?

FLD projections of test image features onto normal vector of decision hyperplane




- LSB trained ensemble classifier with adjusted threshold
- 10% embedding rate
- 5 repetitions of 10-fold-cross validation
- Each test image features projected onto normal vector of decision hyperplane
- Thresholds are medians across all base learners
- Outliers not shown (~12%)



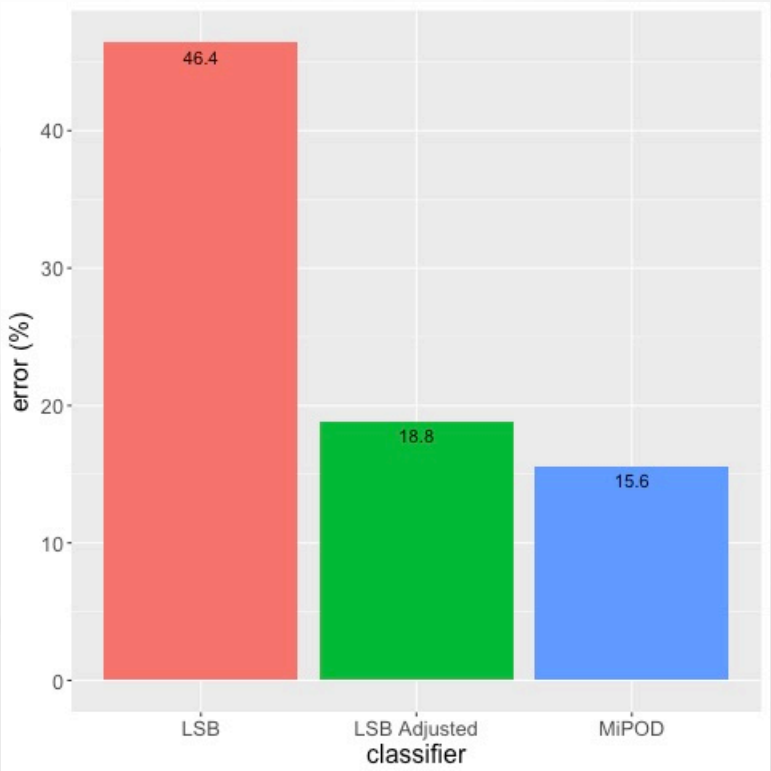
csafe Center for Statistics and Applications in Forensic Evidence

# Terminology

- LSB (MiPOD, S-UNIWARD, WOW) Classifier – an ensemble classifier trained on covers and LSB (MiPOD, S-UNIWARD, WOW) matching data with the standard decision threshold
- LSB Adjusted Classifier – an ensemble classifier trained on covers and LSB matching data with an adjusted decision threshold

# Dataset 1: BOSSbase - 40% Embedding Rate

Average detection error on BOSSbase MiPOD data



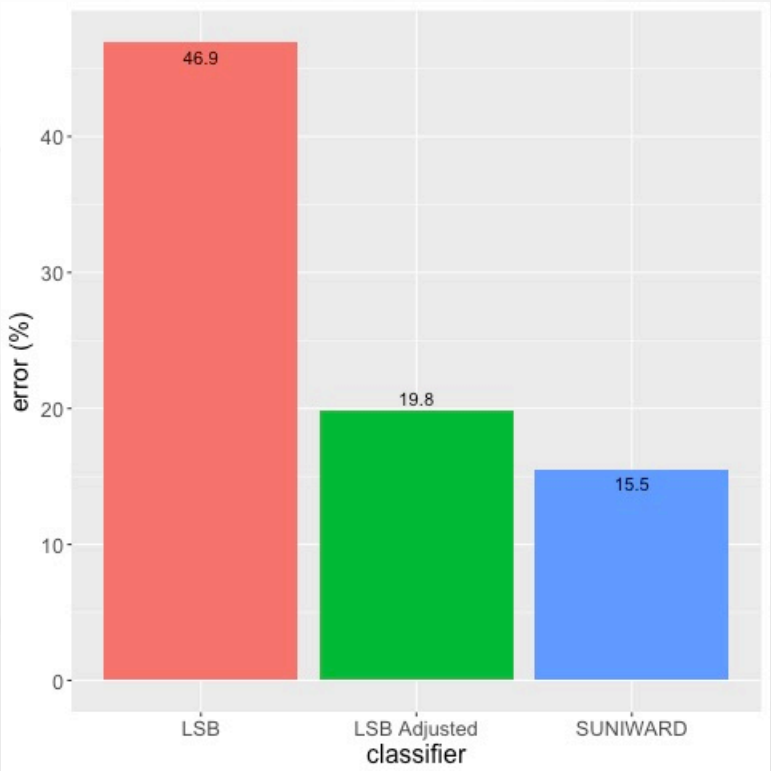

- Testing on MiPOD 40% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - MiPOD classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 40% Embedding Rate

Average detection error on BOSSbase S-UNIWARD data



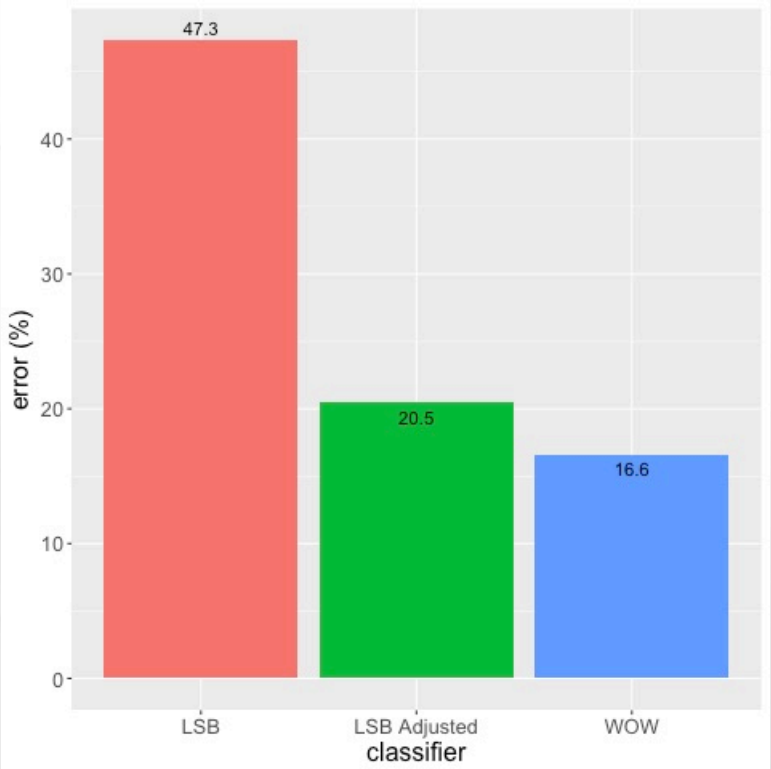

- Testing on S-UNIWARD 40% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 40% Embedding Rate

Average detection error on BOSSbase WOW data



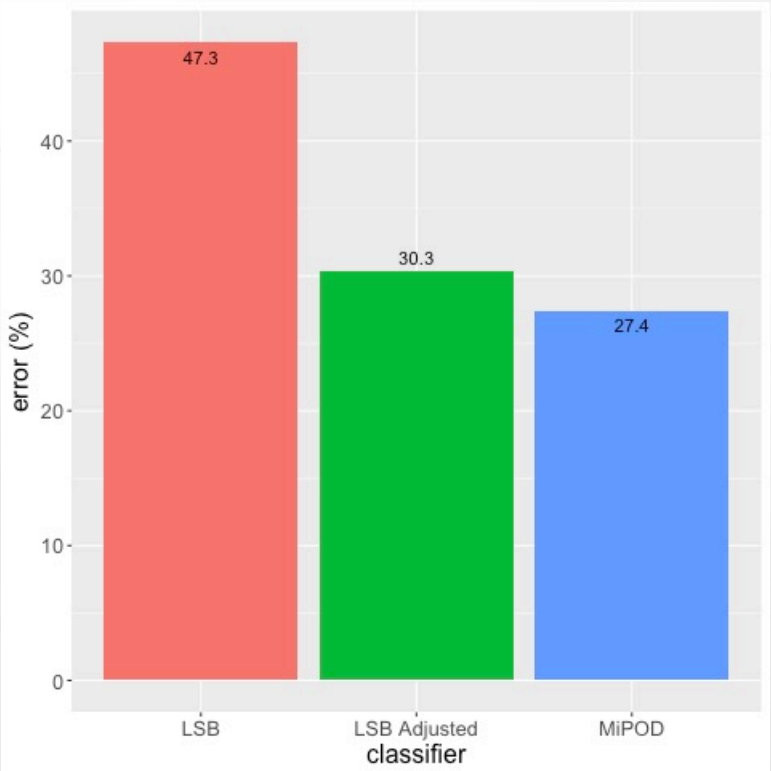

- Testing on WOW 40% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 20% Embedding Rate

Average detection error on BOSSbase MiPOD data



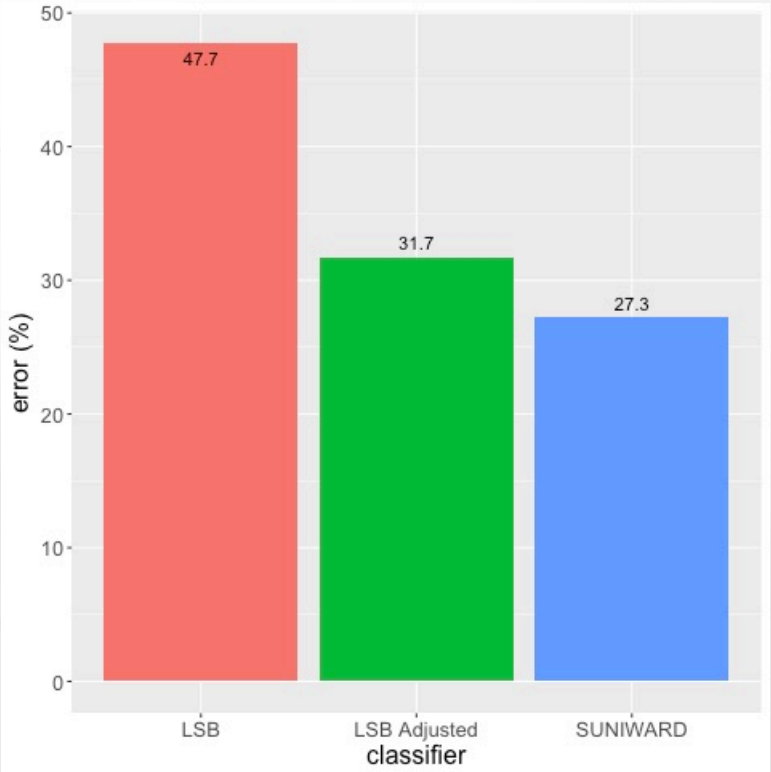

- Testing on MiPOD 20% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - MiPOD classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 20% Embedding Rate

Average detection error on BOSSbase S-UNIWARD data



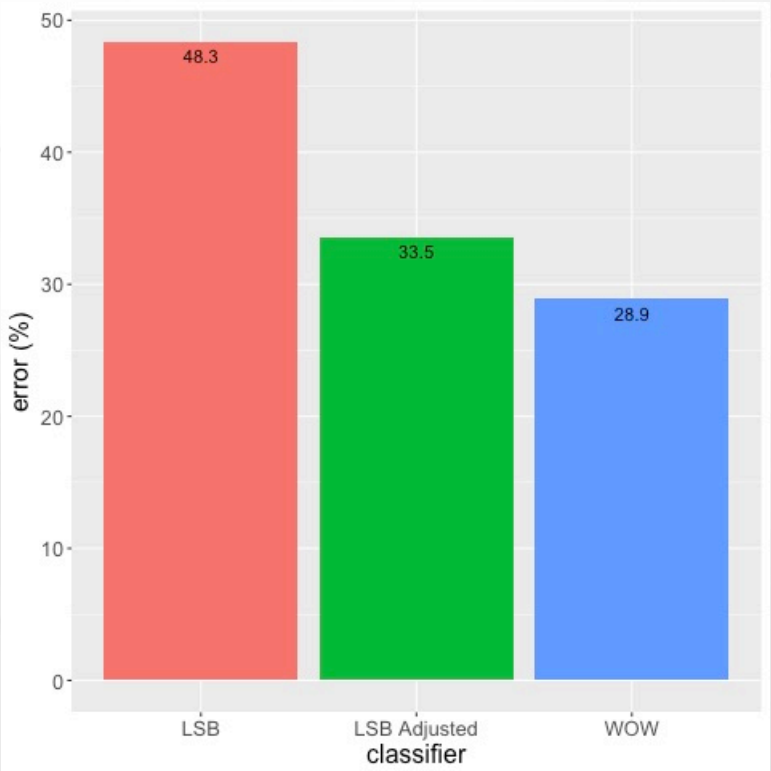

- Testing on S-UNIWARD 20% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - S-UNIWARD classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 20% Embedding Rate

Average detection error on BOSSbase WOW data



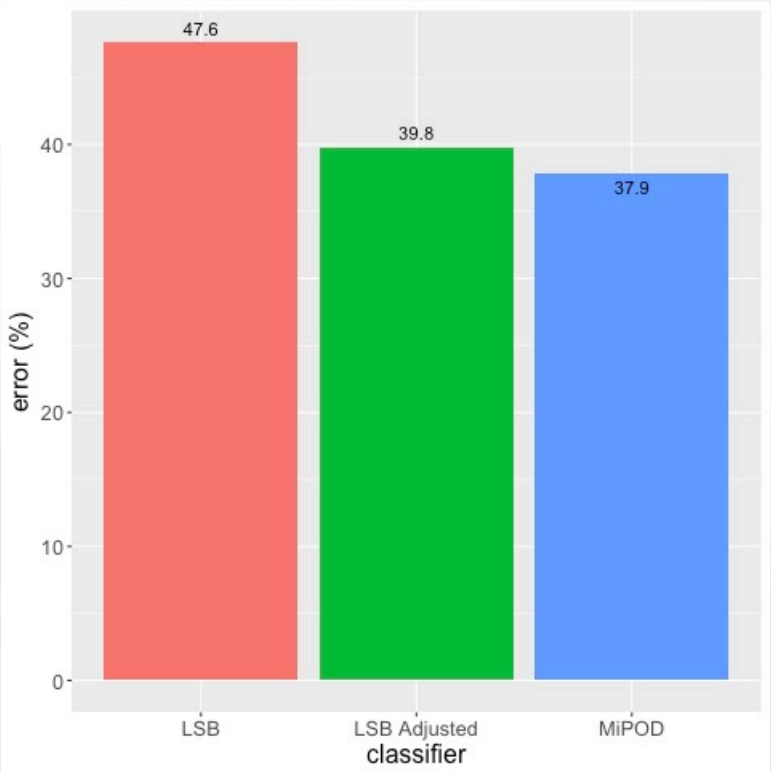

- Testing on WOW 20% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - WOW classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 10% Embedding Rate

Average detection error on BOSSbase MiPOD data



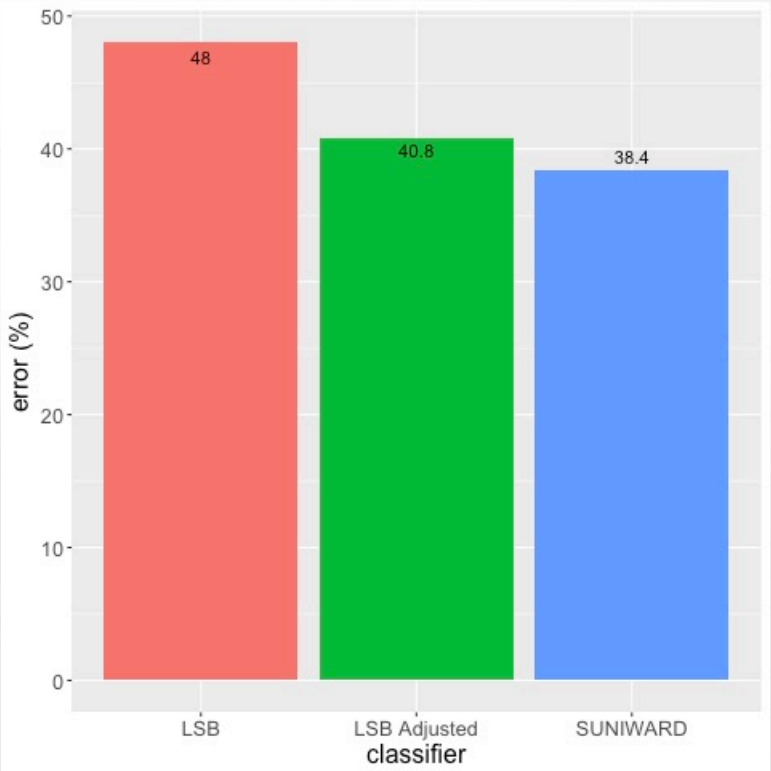

- Testing on MiPOD 10% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - MiPOD classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 10% Embedding Rate

Average detection error on BOSSbase S-UNIWARD data



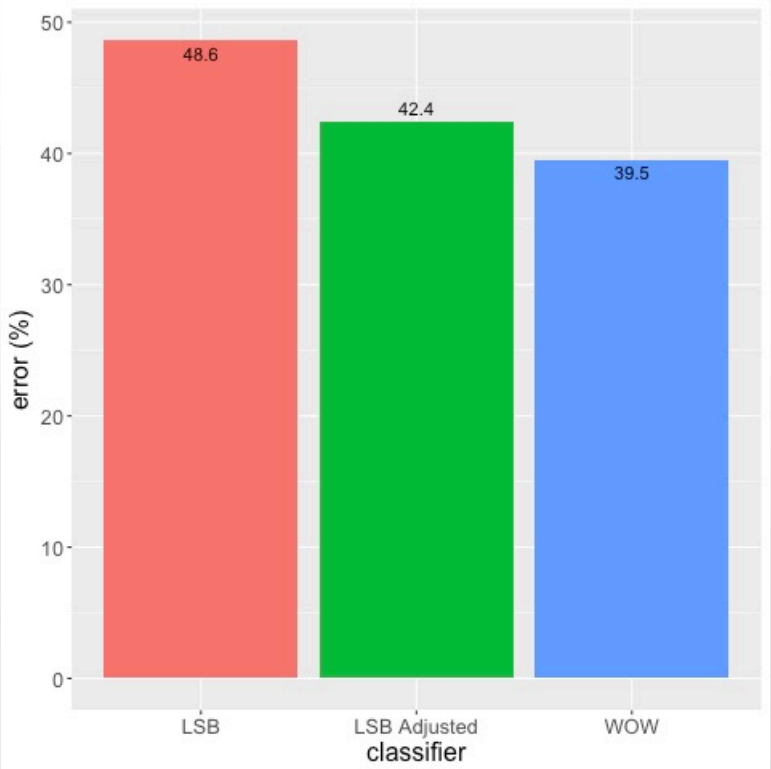

- Testing on S-UNIWARD 10% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - S-UNIWARD classifier uses standard threshold
- LSB adjusted classifier within 2.5% of "best-case" classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 1: BOSSbase - 10% Embedding Rate

Average detection error on BOSSbase WOW data



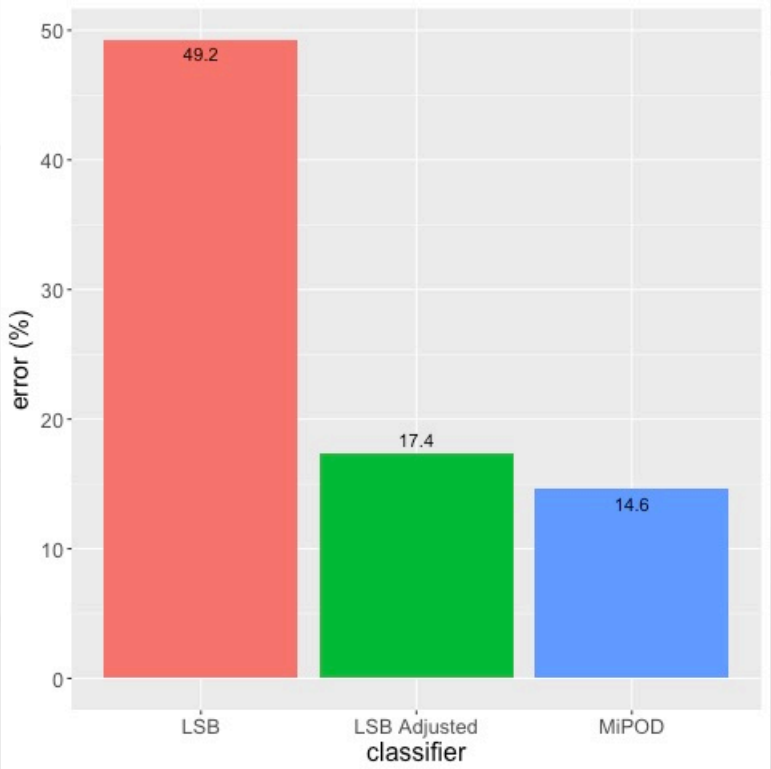

- Testing on WOW 10% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - WOW classifier uses standard threshold
- LSB adjusted classifier within 3% of “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 2: StegoAppDB - 10% Embedding Rate

Average detection error on StegoAppDB MiPOD data



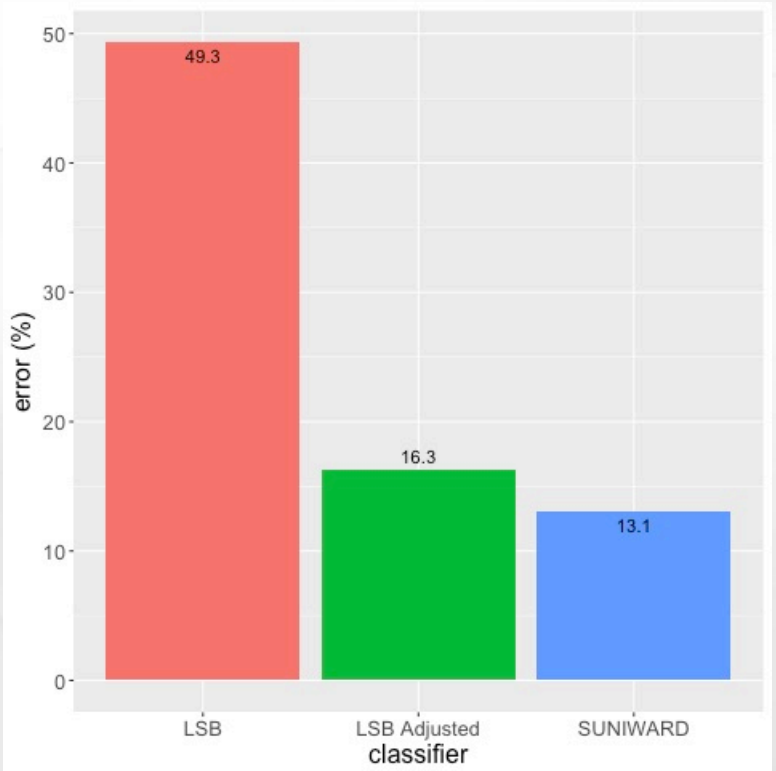

- Testing on MiPOD 10% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - MiPOD classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 2: StegoAppDB - 10% Embedding Rate

Average detection error on StegoAppDB S-UNIWARD data



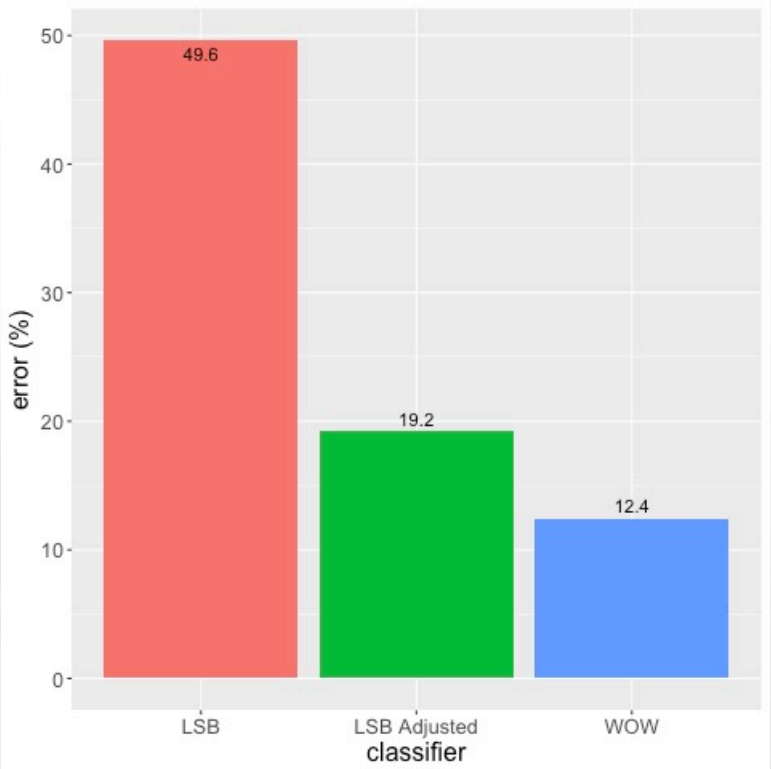

- Testing on S-UNIWARD 10% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - S-UNIWARD classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Dataset 2: StegoAppDB - 10% Embedding Rate

Average detection error on StegoAppDB WOW data




- Testing on WOW 10% embedding rate
  - LSB Adjusted classifier uses adjusted threshold
  - WOW classifier uses standard threshold
  - Training set is 5,000 randomly selected cover-stego pairs and results averaged over 5 repetitions
- LSB adjusted classifier similar results to the “best-case” classifier



csafe
Center for Statistics and Applications in Forensic Evidence

# Conclusions and Future Work

- Conclusions
  - A classifier trained solely on covers and LSB matching can achieve detection error rates close to the "best-case" classifiers when testing on MiPOD, S-UNIWARD, and WOW
- Expand experiments to include more embedding algorithms
- Refine the process of selecting the tuning parameter $\lambda$

# StegoAppDB: Forensic Image Database

- https://data.csafe.iastate.edu/StegoDatabase/

# Acknowledgements


- *This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.*




csafe
Center for Statistics and Applications in Forensic Evidence